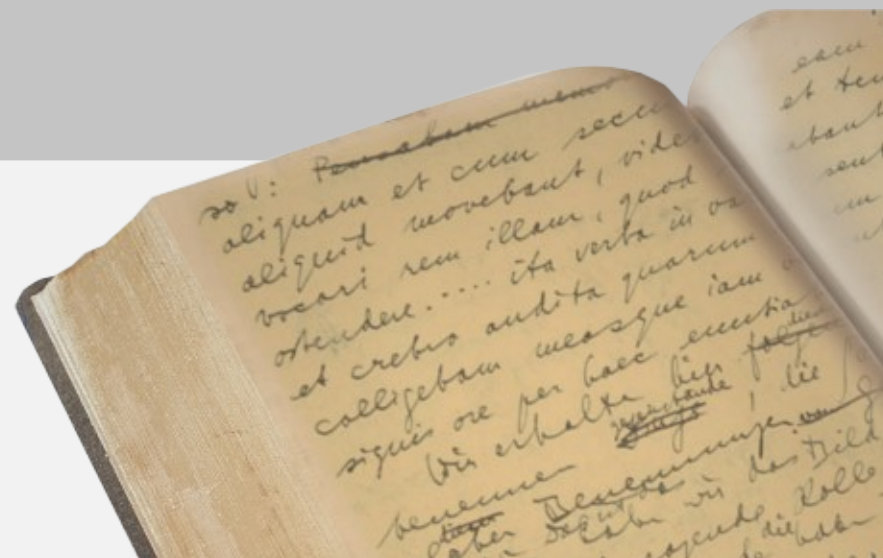


# Towards a Commonwealth of Data, Information and Publications?

## Librarian Agendas at the End of the Gutenberg Paranthesis

*Prof. Dr. Stefan Gradmann (KU Leuven)*

National Library-Ex Libris Annual Meeting, Singapore, 21/08/2013



- Container → Content → Context: Libraries approaching the End of the Gutenberg Galaxis and the emerging Linked Data Web
- Linked Data enabling Research with content based and context driven services
- Required Cultural Changes: terms/thinking to get rid of
- Strategy issues: what about the question mark?



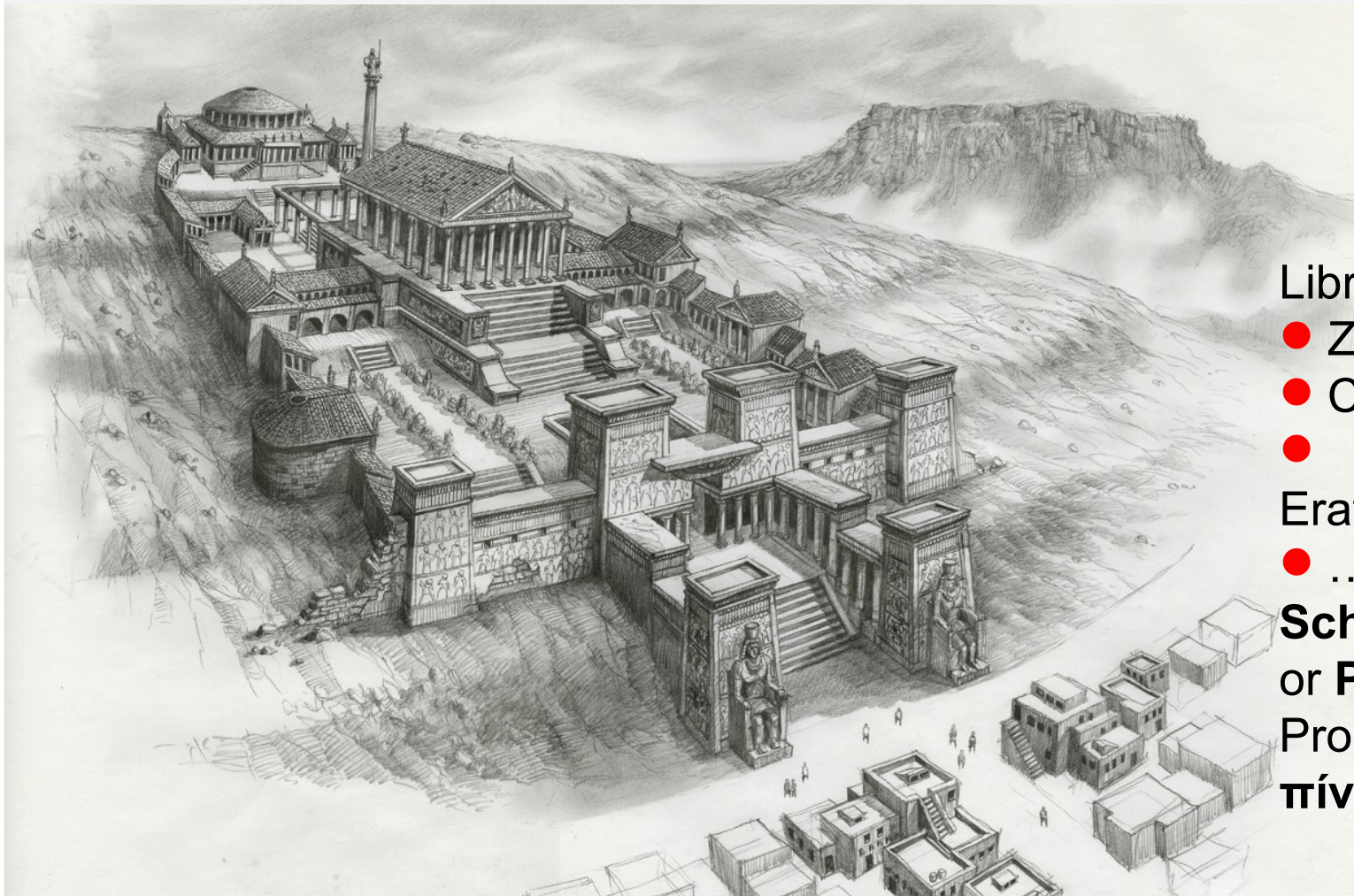
# Container → Content → Context

## Libraries approaching the end of the Gutenberg Paranthesis



# Long before the Parenthesis: Alexandria

KU LEUVEN



Librarians:

- Zenodotus
- Callimachus
- 

Erathosthenes

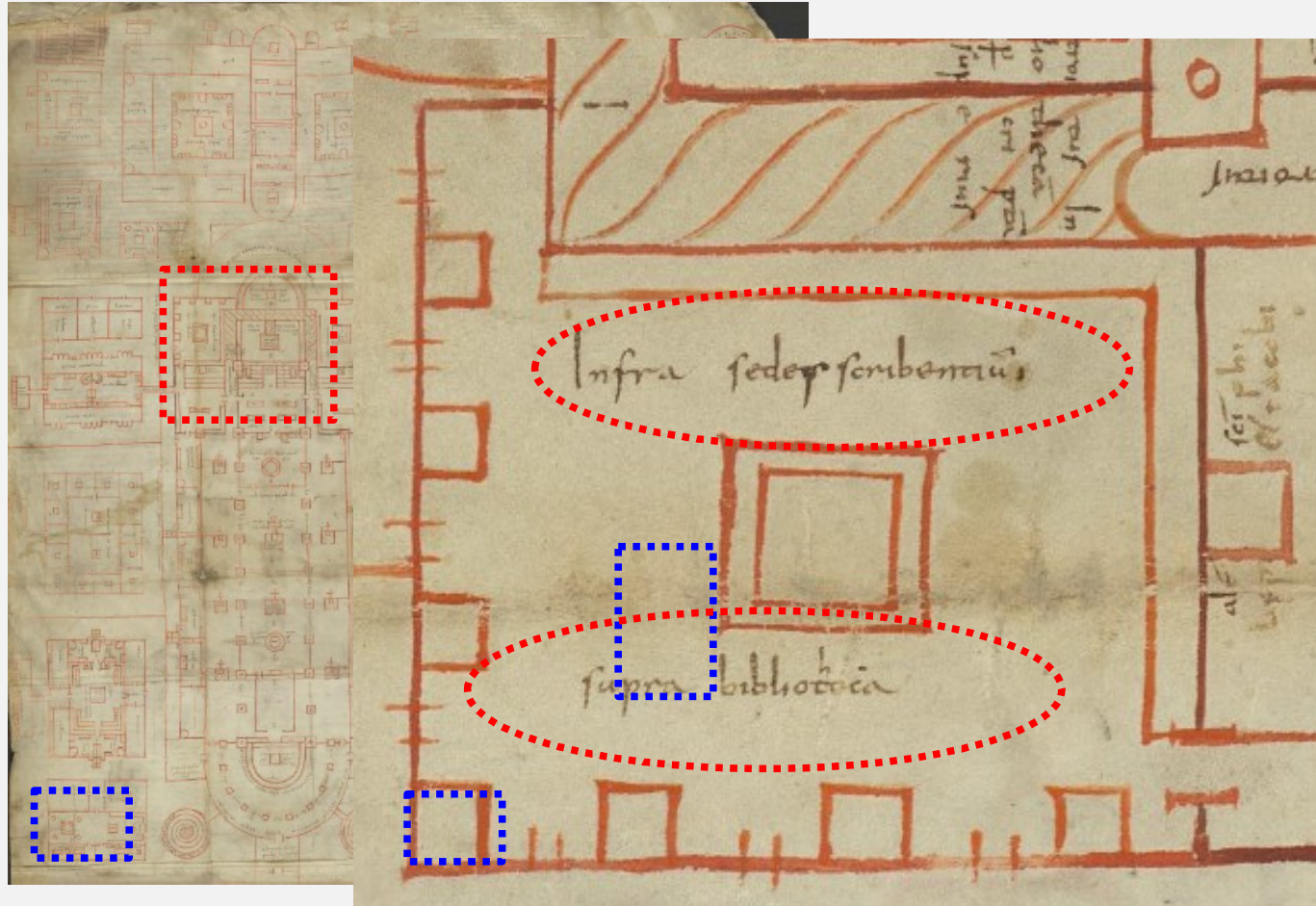
- ...

**Scholars** and /  
or **Poets**  
Producing  
**τίνakes**





# Before the Parenthesis: St. Gall

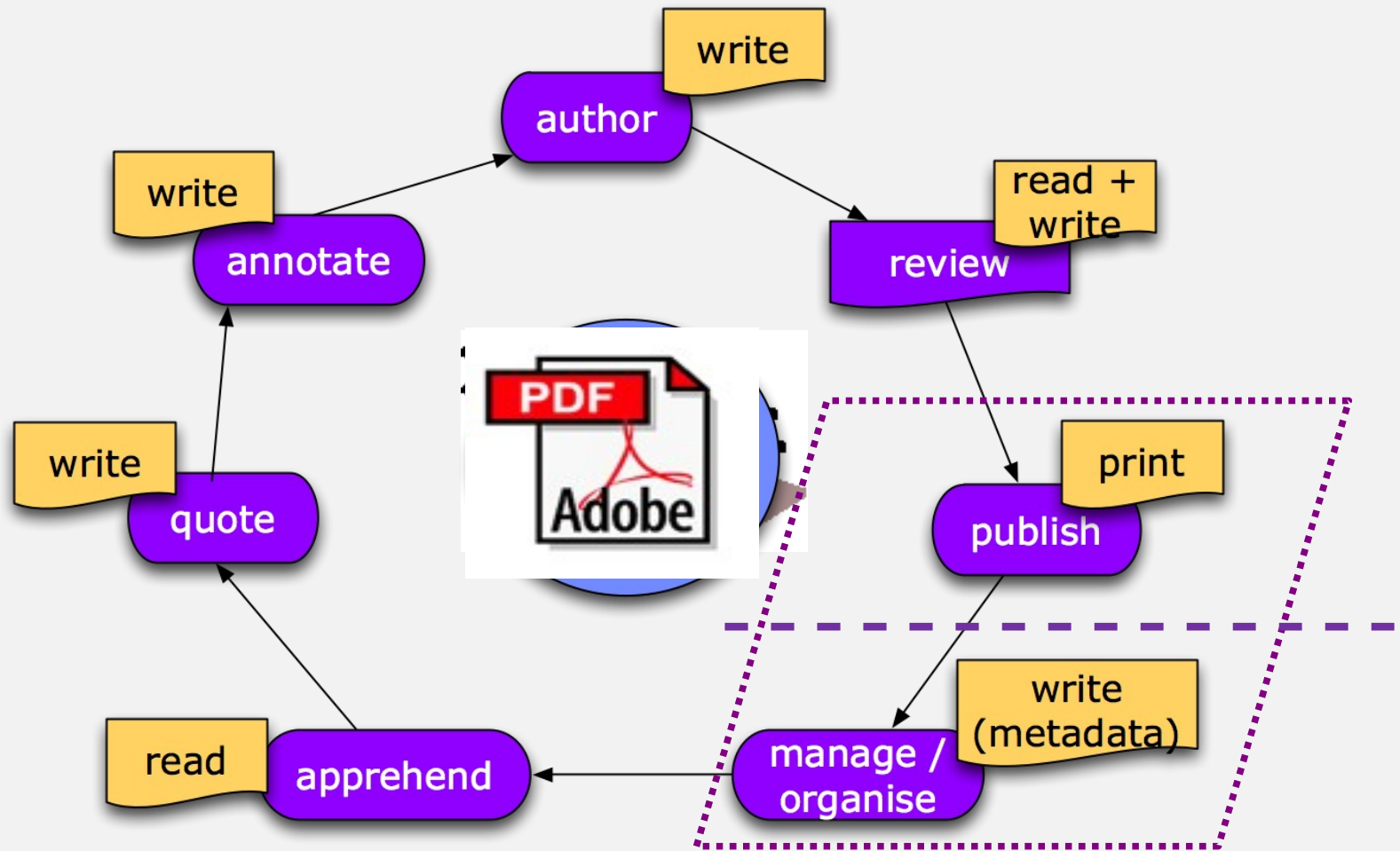


# The Gutenberg Parenthesis Opens ...

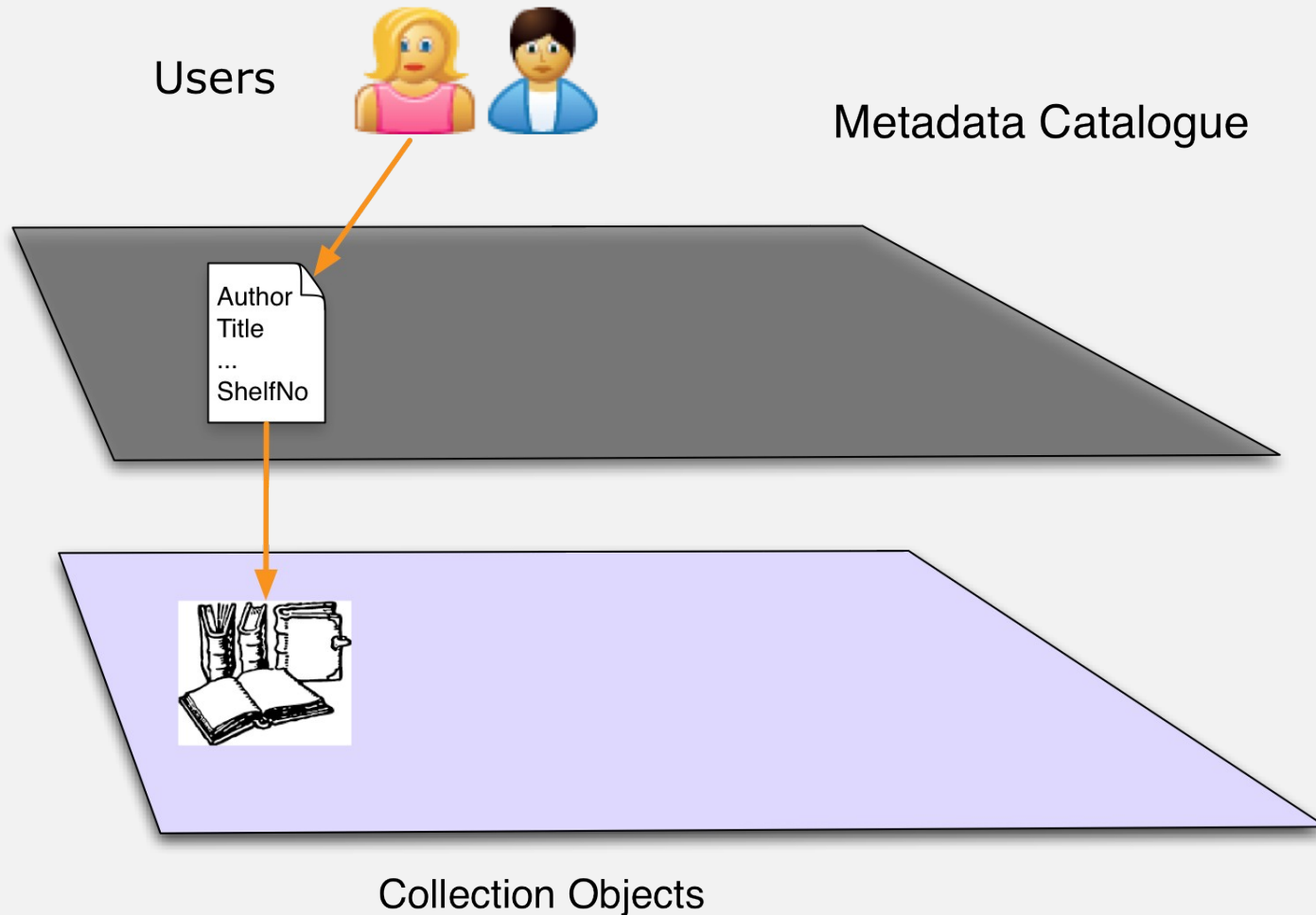
## Dissociation of container and content in the print paradigm



# Dissociation of Roles in the Gutenberg galaxy

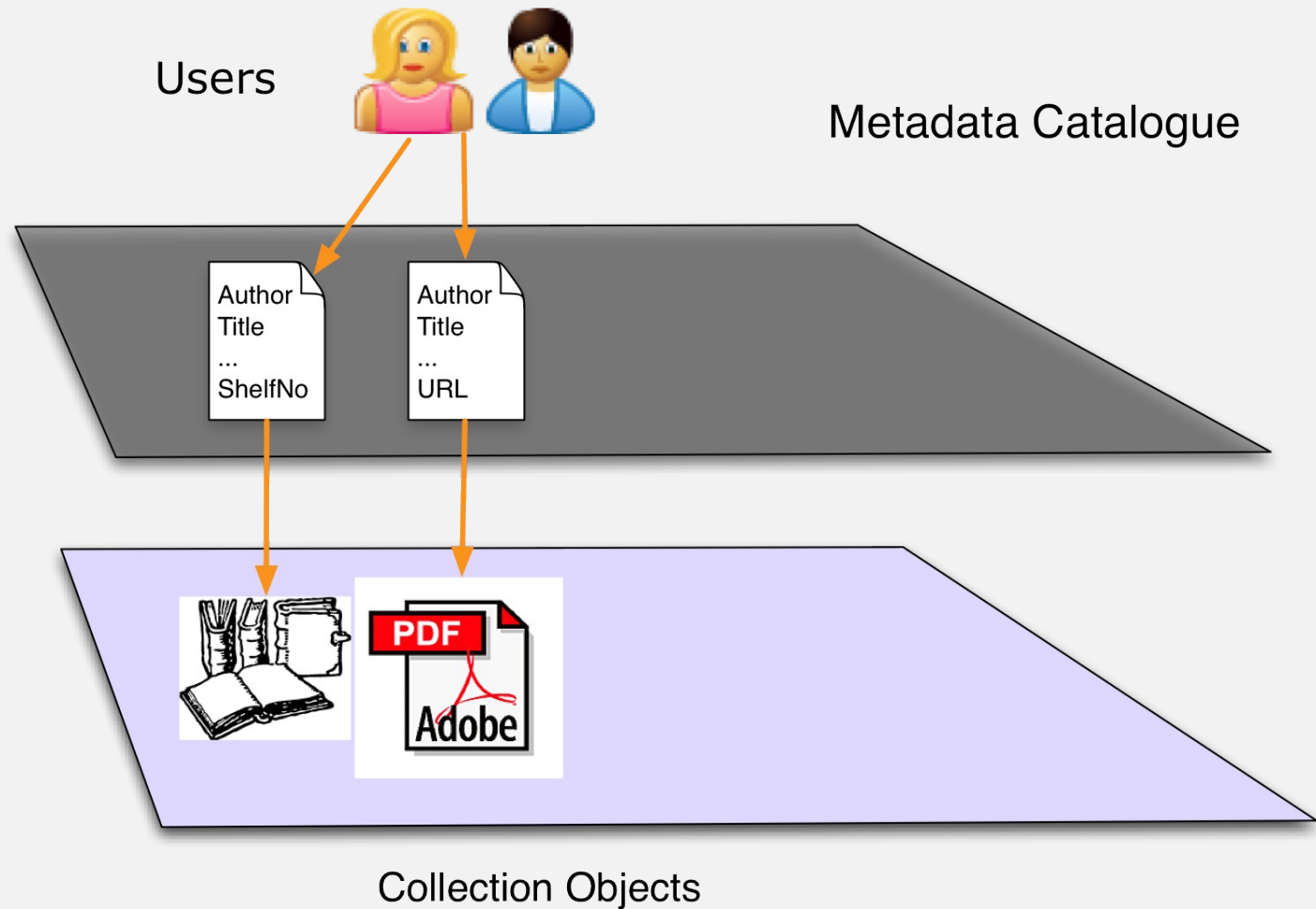


# Catalogue Based Library Functional Axioms (1)

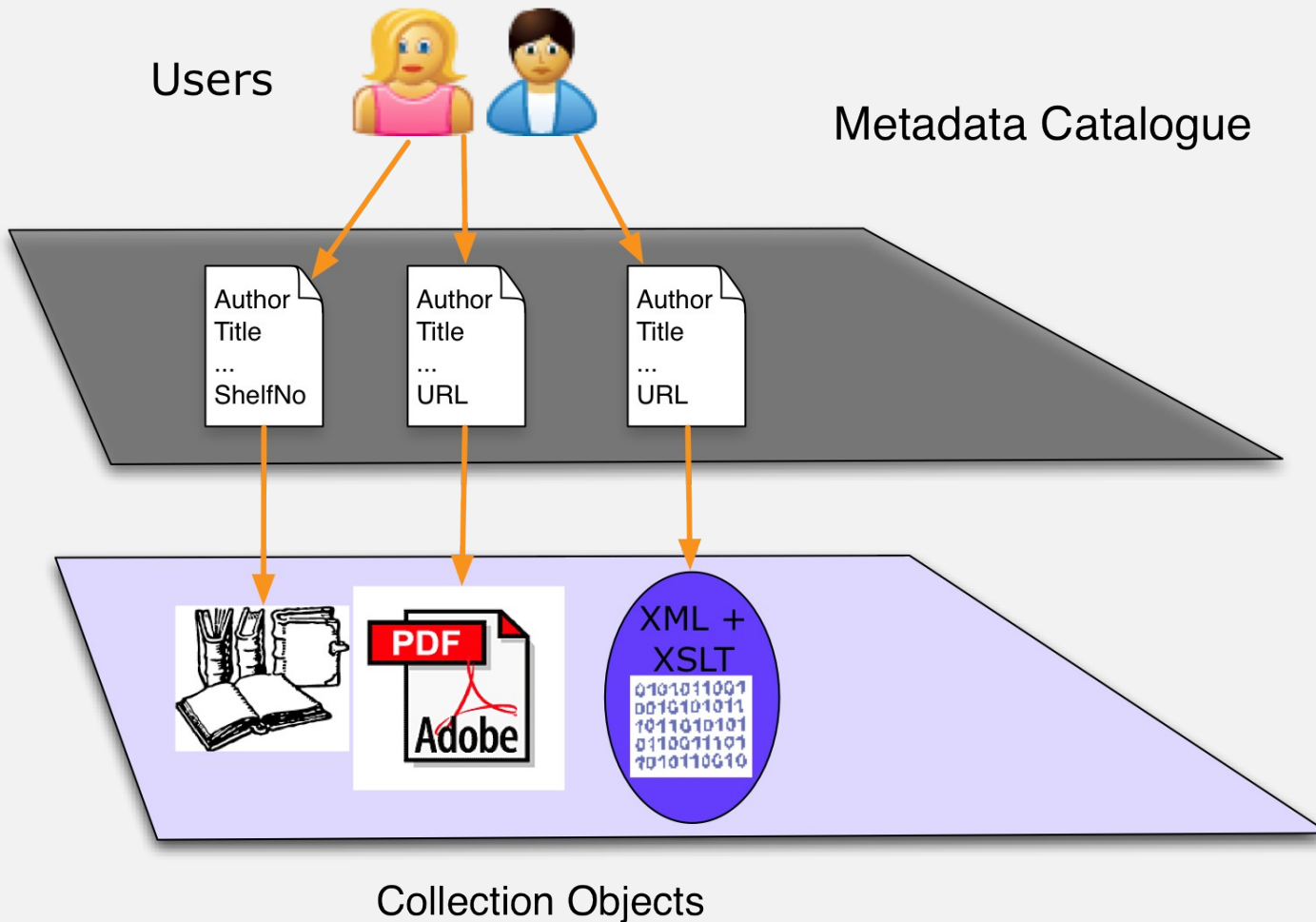




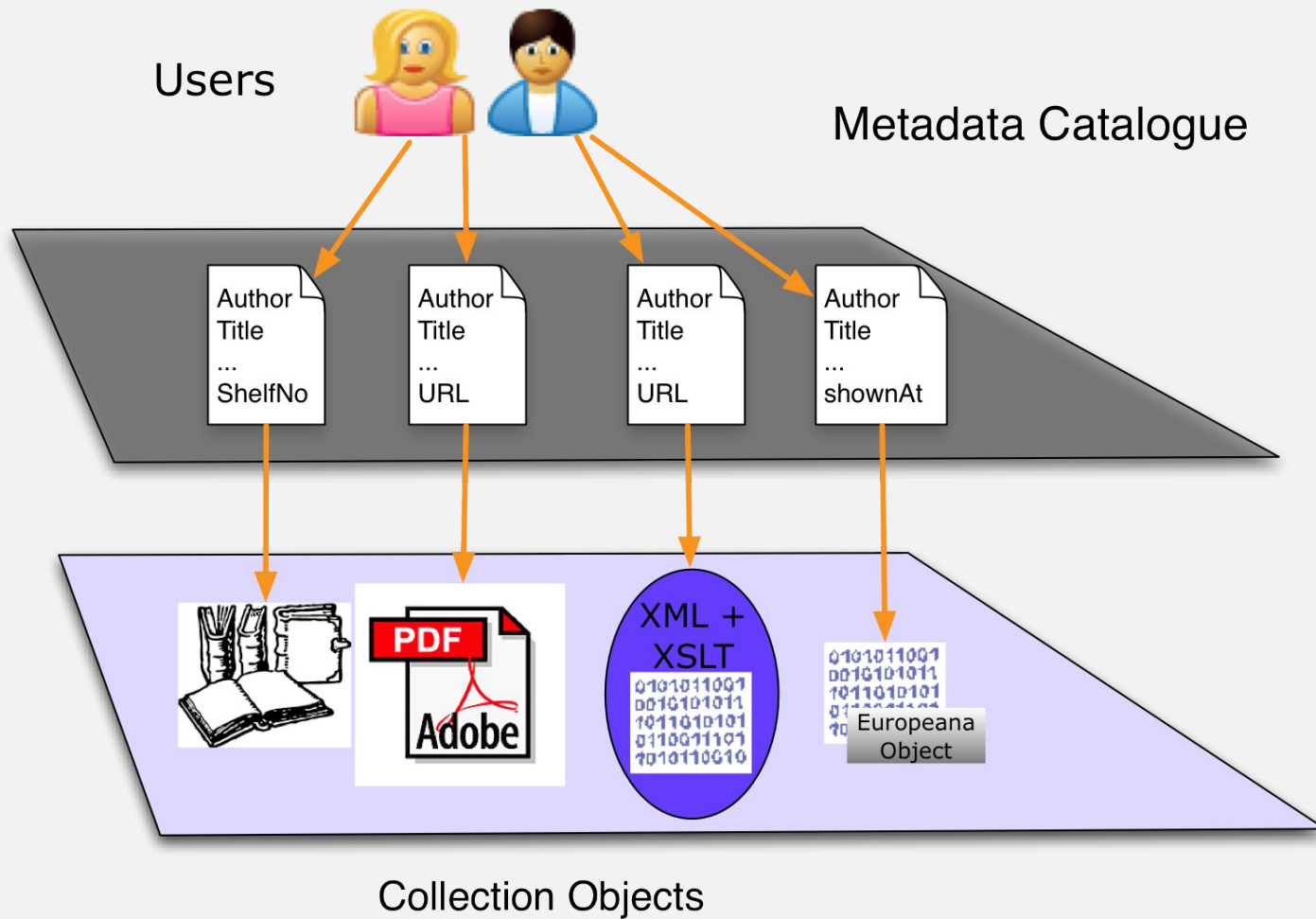
# Catalogue Based Library Functional Axioms (2)



# Catalogue Based Library Functional Axioms (3)



# Catalogue Based Library Functional Axioms (4)



- **Mediating access** to information objects via catalogues
- **Mediating links** as pointers from metadata to objects
- Objects are part of a library **collection**
  - An object to be used within a library typically is part of this library's collection
- Internal processing logic: focus on
  - objects as information **containers**,
  - not so much on the **content** of these containers
  - and accordingly **cataloguing** is focussed on **container attributes**
- Functional macro-primitives are **ingestion, storage, description and retrieval** of information containers



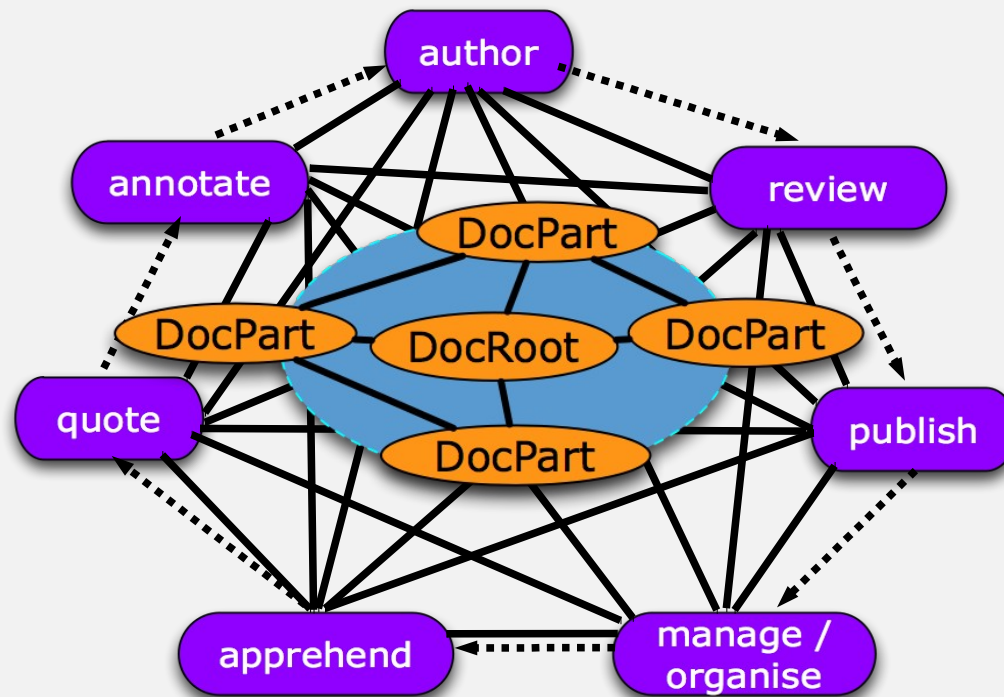
# ... and closes again

## The end of the print paradigm





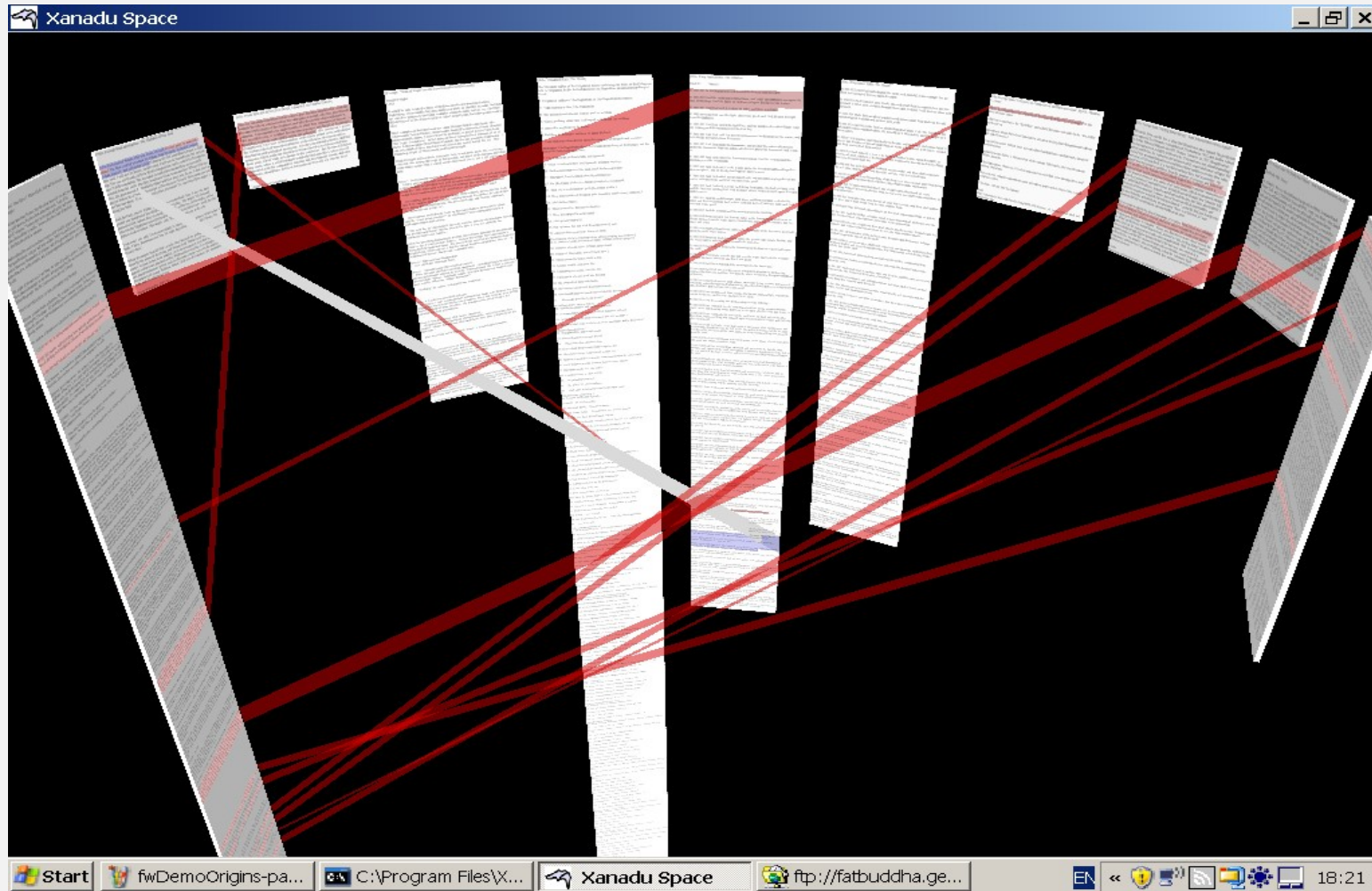
# Web Based Scholarly Working Continuum ... ... a triple paradigm shift: Beyond Documents

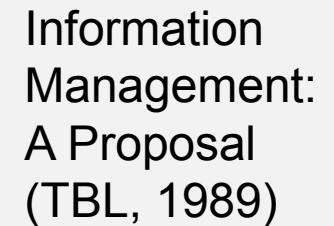


- Decreasing functional determination by traditional cultural techniques
- Disintegration of the linear / circular functional paradigm
- Erosion of the monolithic document notion in hypertext paradigms



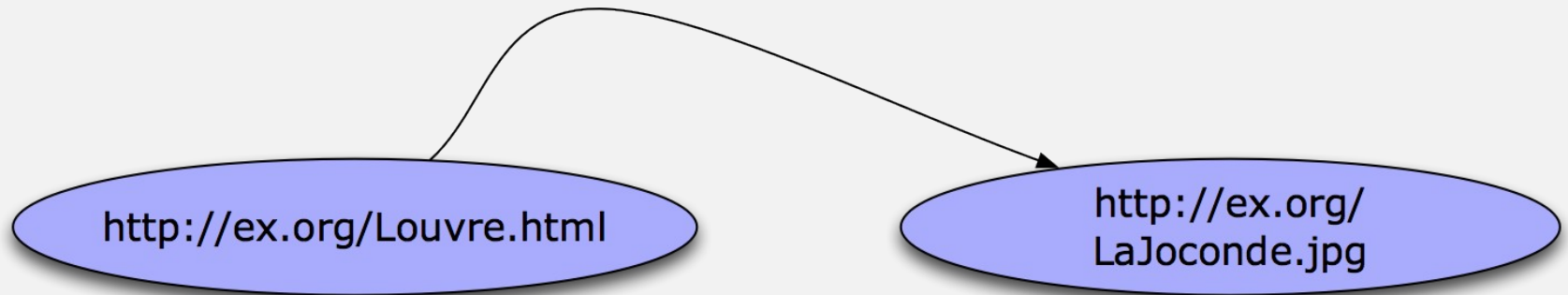
# Ted Nelson's Xanadu: radicalised Hypertext ...





- ... twice
- extended:
  - in syntax
  - in scope

# Resources and Links in the Document Web



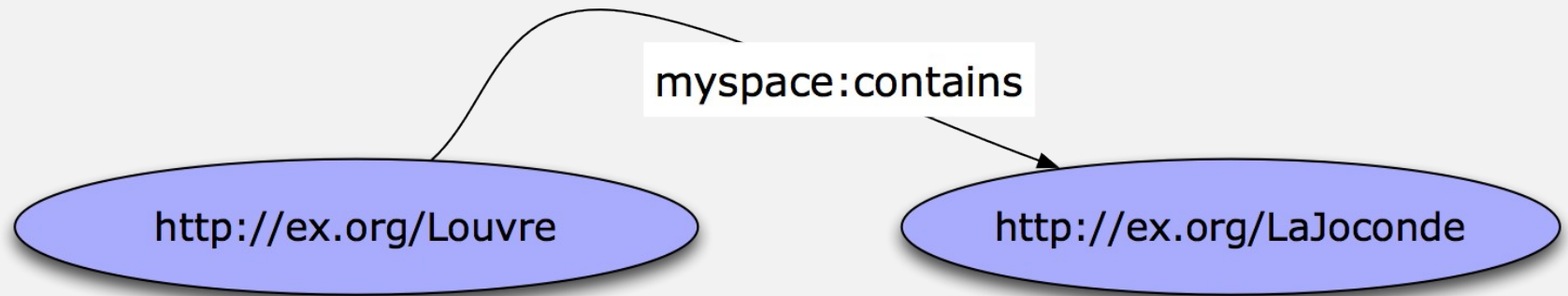
```
<a href="http://ex.org/LaJoconde.jpg">b</a>
```

- We have HTTP URIs to identify resources and links between them – but we are missing a few things!
- What kinds of resources are 'Louvre.html' and 'LaJoconde.jpg'?
  - A machine cannot tell.
  - Humans can: we recognize implied context!
- How exactly do they relate to each other?
  - A machine cannot tell.
  - Humans can: again we recognize implied context!



# Syntactically Extending the Document Web (1)

- We add a syntax for making statements on resources: RDF triples

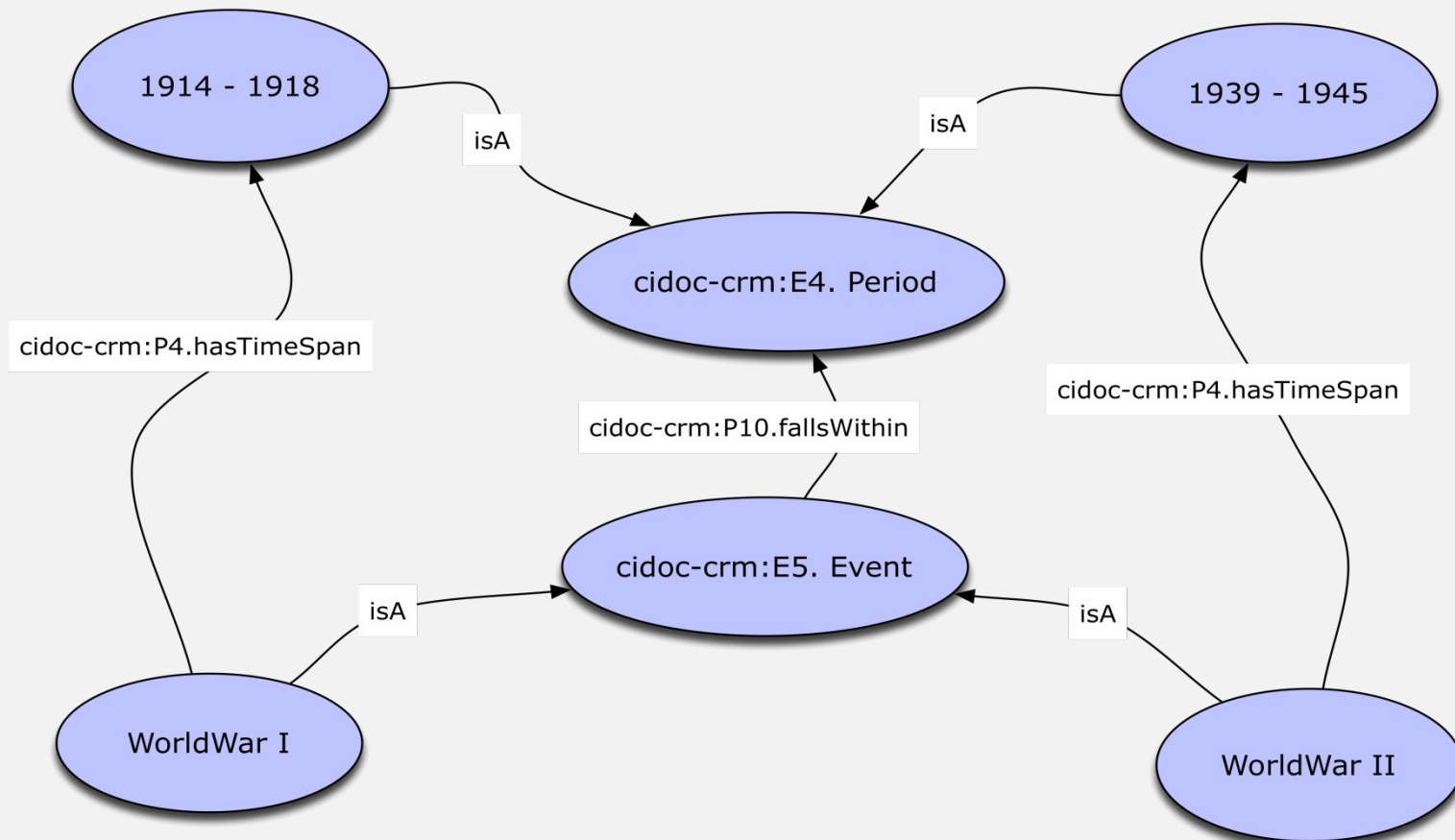


- We add a schema language (RDFS) with elements such as
  - classes ('chair' as instance of chairs),
  - hierarchies of classes and properties (chairs are a subclass of furniture, 'teaches' is a sub-property of 'communicates')
  - inheritance (communication based on language → teaching also is)
  - support for basic inferencing, deterministic logical



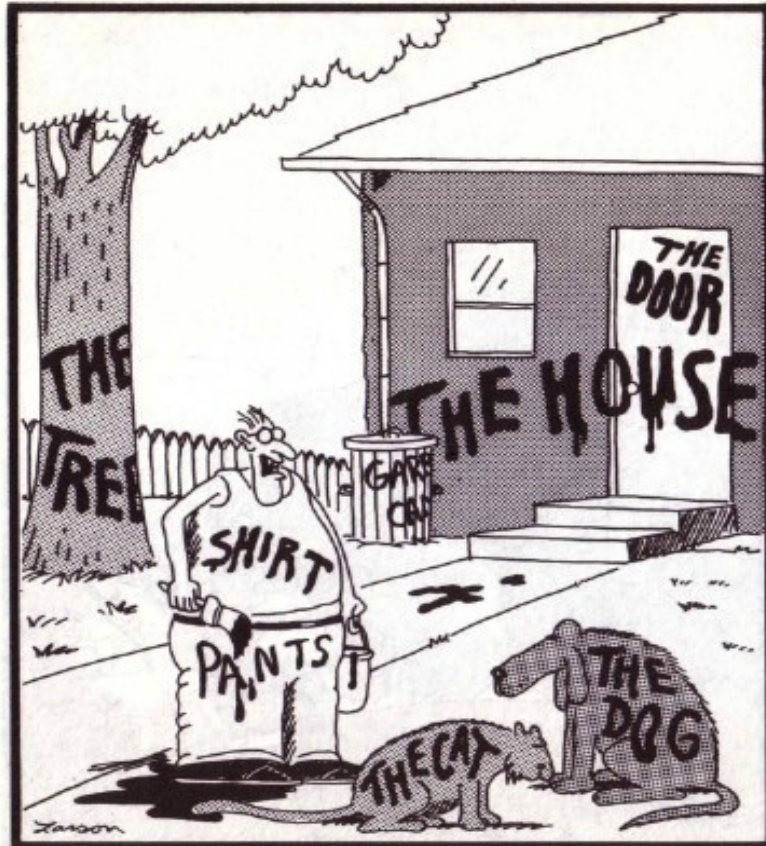
# Syntactically Extending the Document Web: RDF (2)

- And thus are able to establish structures in triple aggregations resulting in lightweight domain ontologies:



# Extending the Web in Scope: The Web of Things ... (slightly Mistaken)

KU LEUVEN



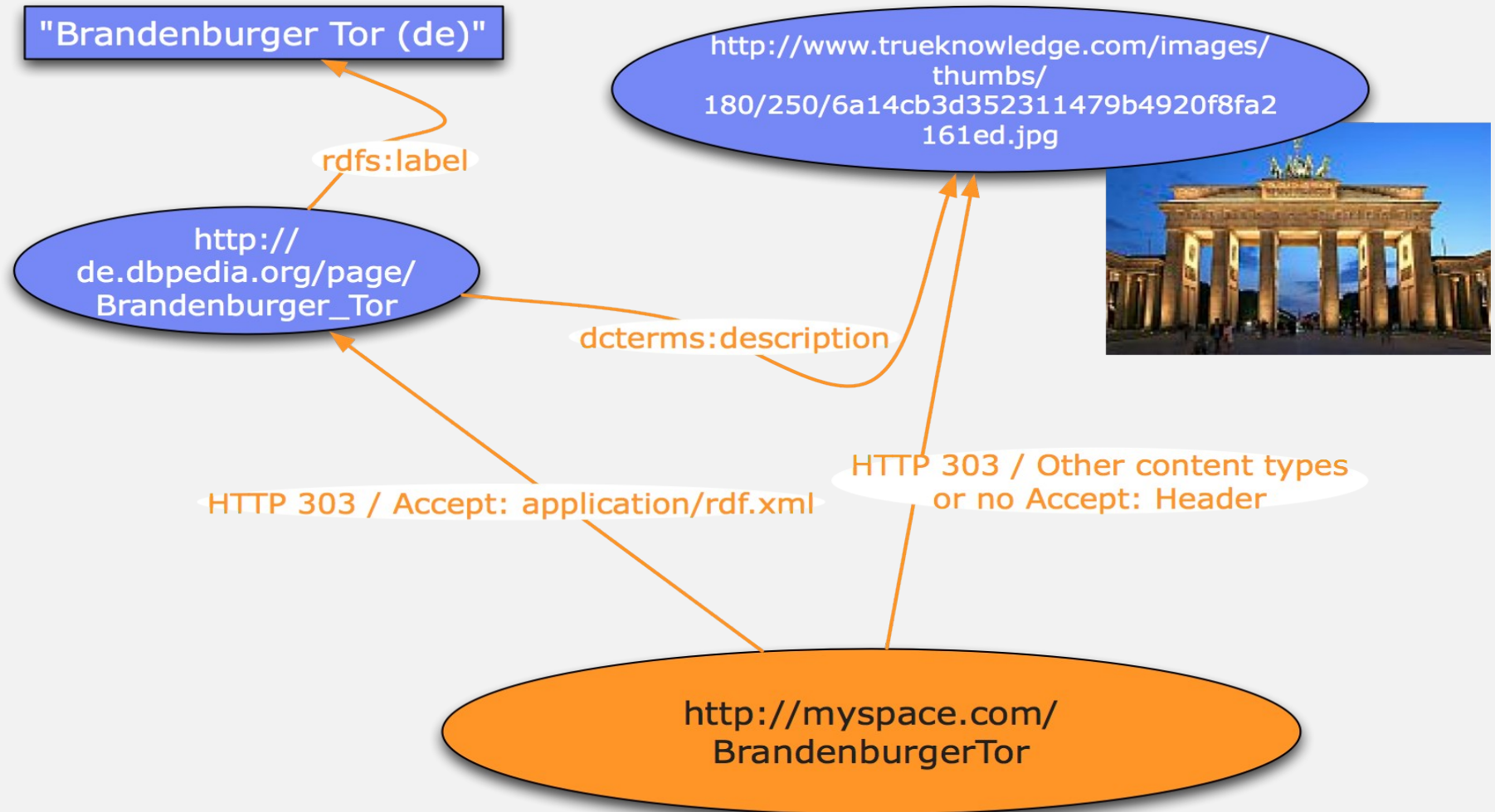
"Now! ... That should clear up  
a few things around here!"

What's **wrong**  
with this picture?

Taken from Ronald Carpentier's  
Blog at  
<http://carpentier.wordpress.com/2007/08/08/1-2-3/>



# ... and the Way we extend the Web in scope to make it a 'Web of Things'



# And we get ... Linked Data

## Linked Data essentials

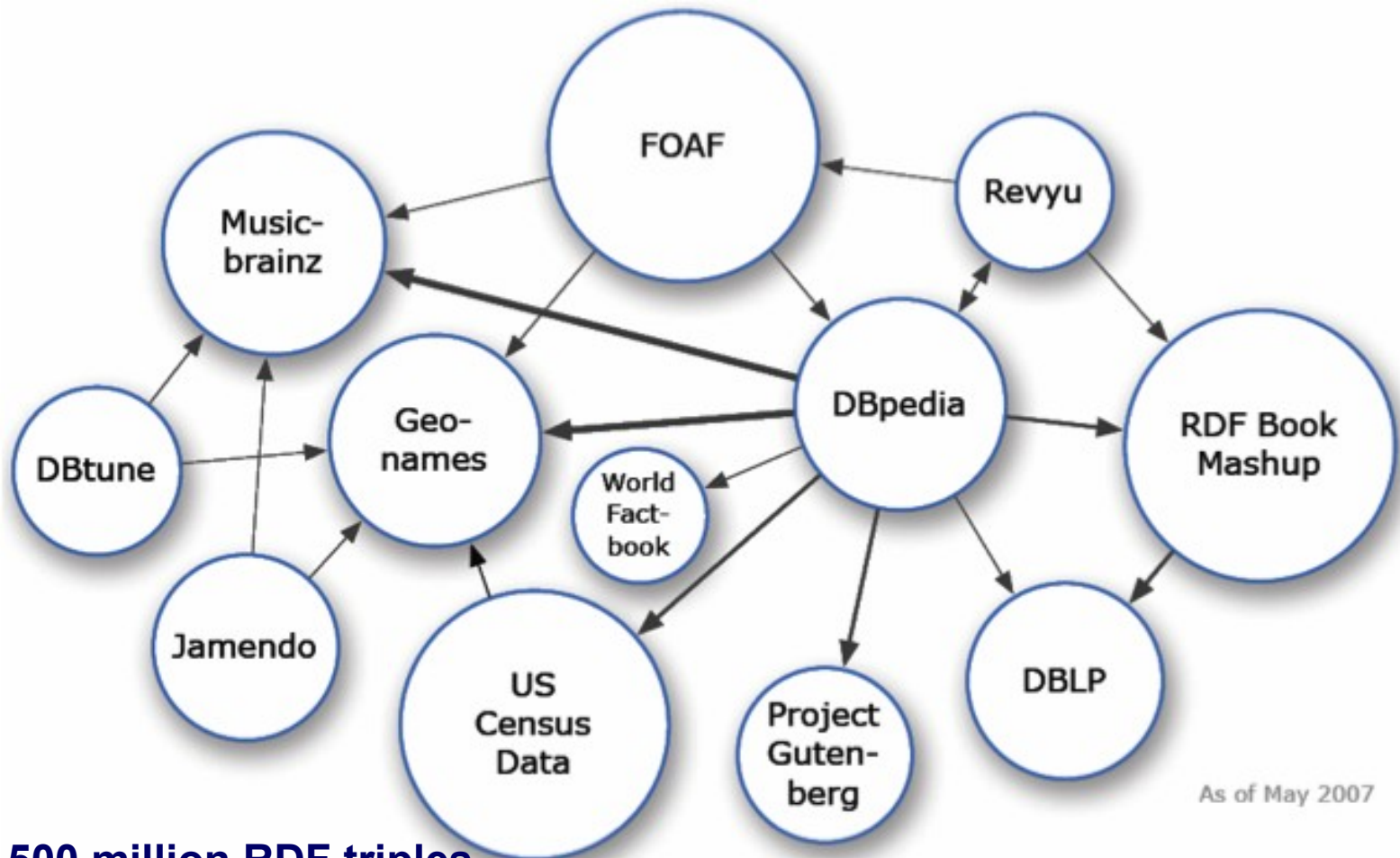
1. Use URIs      Standard Identifiers
  2. Use HTTP URIs      Standard Pointers
  3. Serve useful information using SPARQL, RDF standards
  4. Mention URIs of related objects      Standards for Queries and Statements
- Link to Context

Copyright © 2008 W3C (MIT, ERCIM, Keio)

[http://www.w3.org/2008/Talks/0617-lod-tbl/#\(4\)](http://www.w3.org/2008/Talks/0617-lod-tbl/#(4))



# A few Bubbles: 5/2007



Over 500 million RDF triples

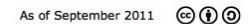
Around 120,000 RDF links between data sources

© Richard Cyganiak

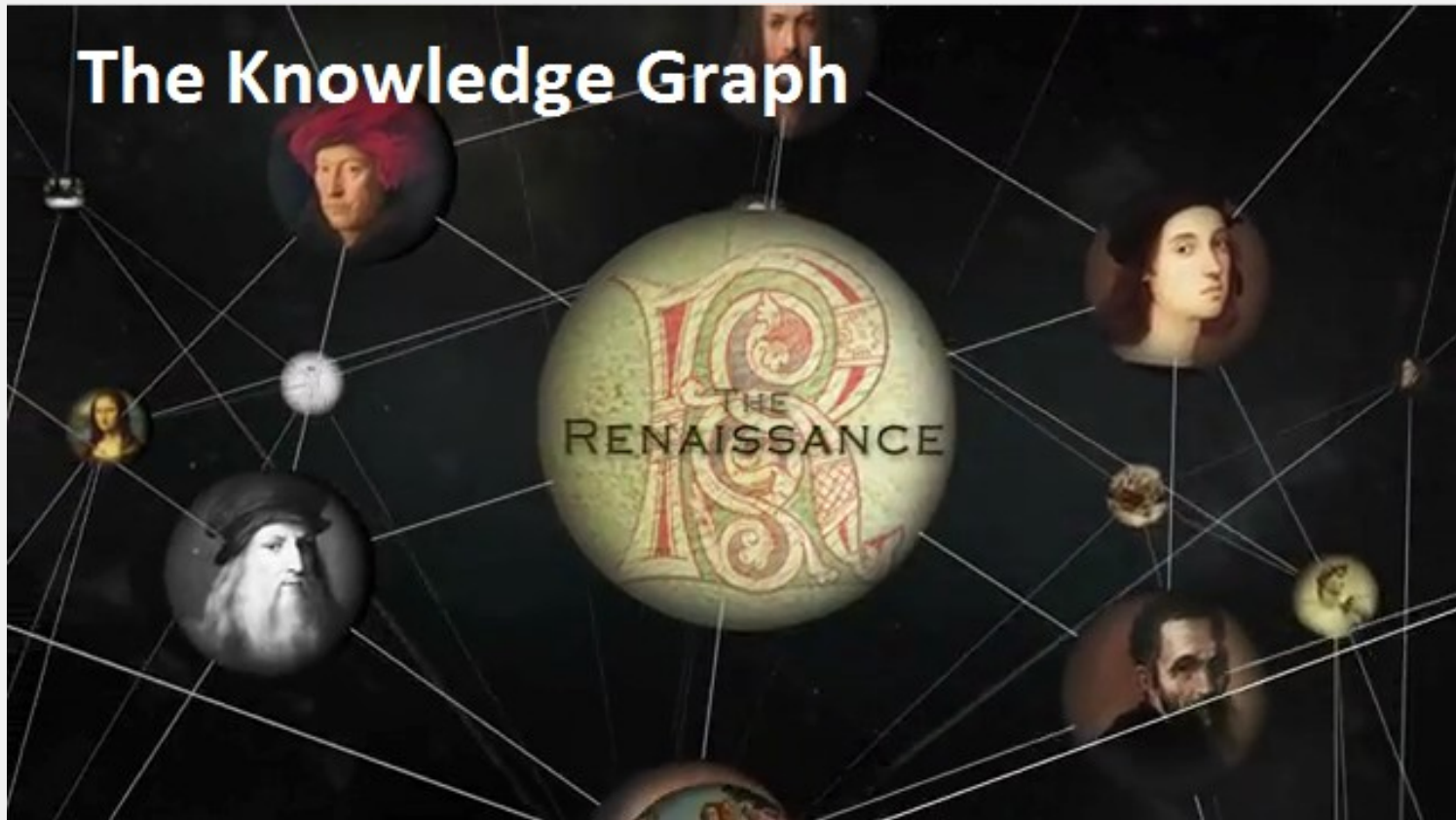




**KU LEUVEN**



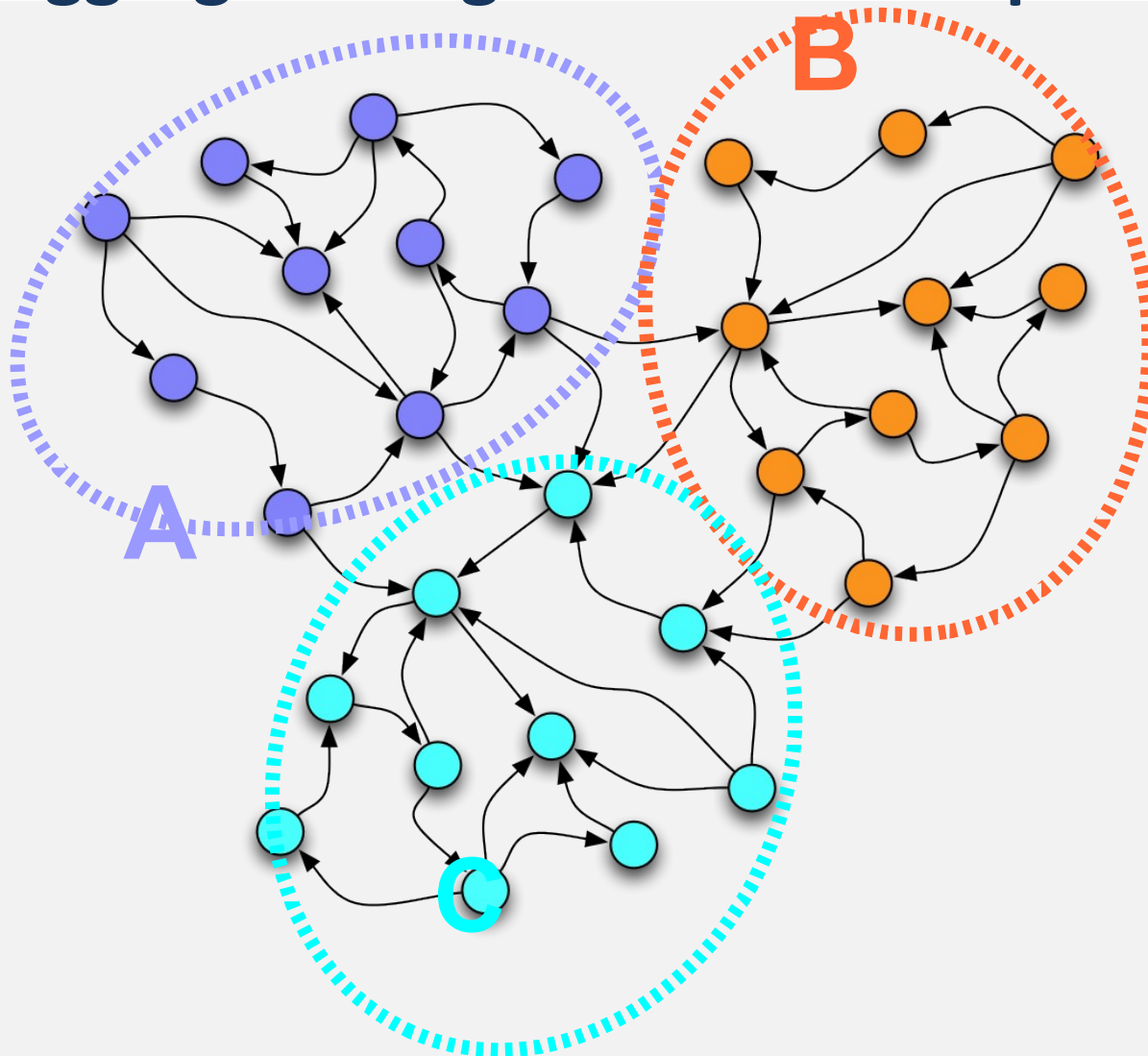
# Google entering the Floor





# Modelling Object Representations as RDF

## Aggregations generates new questions ...

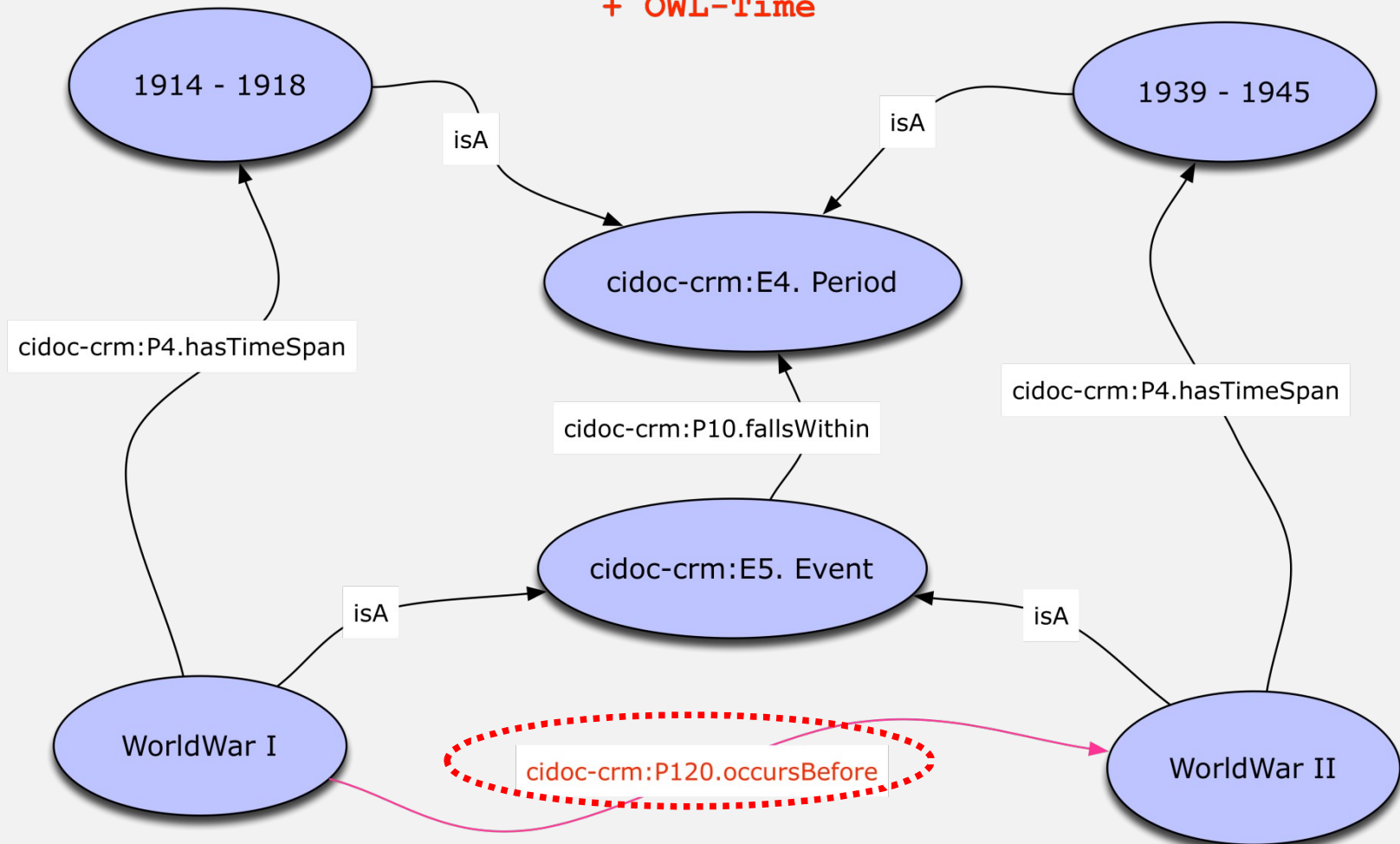


- Where do resource aggregations 'start'? Where do they 'end'?
- And what constitutes document boundaries??
- And which node was connected to which one at a given time???  
→ Provenance, Versioning, Authorisation: Named Graphs, Ways of making Meta-Statements



# ... and new opportunities: Triple Sets and 'Reasoning'

+ OWL-Time



# ... based on 'Documents' as Aggregations of RDF-Triples (1)

## Assertion

NG\_000007.  
3:g.70628G>A

has  
frequency

0.25%

Condition

Sardinian

Provenance

Giardine et al





# The use of Inferences

KU LEUVEN

van Haagen HHHBM, 't Hoen PAC, Botelho Bovo A, de Morrée A, van Mulligen EM, et al. (2009) Novel Protein-Protein Interactions Inferred from Literature Context. PLoS ONE 4(11): e7894. doi:10.1371/journal.pone.0007894 / Example provided by Jan Velterop

 **PLoS one**  
accelerating the publication of peer-reviewed science

Login | Create Account | Feedback

Search articles... **GO** Advanced Search

**Browse** **RSS**

Home Browse Articles About For Readers For Authors and Reviewers Journals Hubs PLoS.org

**RESEARCH ARTICLE** **OPEN ACCESS**

**Novel Protein-Protein Interactions Inferred from Literature Context**

Article Metrics Related Content Comments: 0

**Herman H. H. B. M. van Haagen<sup>1\*</sup>, Peter A. C. 't Hoen<sup>1</sup>, Alessandro Botelho Bovo<sup>2</sup>, Antoine de Morrée<sup>1</sup>, Erik M. van Mulligen<sup>1</sup>, Christine Chichester<sup>1</sup>, Jan A. Kors<sup>1</sup>, Johan T. den Dunnen<sup>1</sup>, Gert-Jan B. van Ommen<sup>1</sup>, Silvère M. van der Maarel<sup>1</sup>, Vinícius Medina Kern<sup>2</sup>, Barend Mons<sup>1</sup>, Martijn J. Schuemie<sup>1</sup>**

**1** BioSemantics Association, Department of Human Genetics, Leiden University Medical Center, Leiden, and Department of Medical Informatics, Erasmus University Medical Center, Rotterdam, The Netherlands, **2** Post-Graduate Program in Knowledge Engineering and Management (EGC), Federal University of Santa Catarina (UFSC), Florianópolis, Brazil

**Abstract** [Top](#)

We have developed a method that predicts Protein-Protein Interactions (PPIs) based on the similarity of the context in which proteins appear in literature. This method outperforms previously developed PPI prediction algorithms that rely on the conjunction of two protein names in MEDLINE abstracts. We show

To add a note, highlight some text. [Hide notes](#)  
Make a general comment

**Jump to**  
[Abstract](#)  
[Introduction](#)  
[Results](#)  
[Discussion](#)  
[Methods](#)  
[Supporting Information](#)  
[Author Contributions](#)  
[References](#)

**Download:** [PDF](#) | [Citation](#) | [XML](#)  
[Print article](#)

**Metrics** ⓘ  
**Total Article Views:** **673**  
**Average Rating** (0 User Ratings)  
☆☆☆☆ [See all categories](#)  
[Rate This Article](#) **More**

**Related Content**  
**Related Subject Categories**  
[Molecular Biology](#), [Computer Science](#), [Computational Biology](#)  
**Related Articles on the Web**  
[Google Scholar](#)  
[PubMed](#) **More**

**Share this Article** ⓘ  
Bookmark:        
[Email this article](#)



# Data = Publication

- Distinction data vs. publication will get **increasingly obsolete** in semantic publishing environments ...
- ... at least in the STM sector.
- The move into semantic publication will be **much slower in the SSH** because of
  - fuzzy and unstable **terminology**
  - fuzzy **linking semantics** hard to formalise consistently
  - close relation between **complex document formats** and **scholarly discourse**
- Current examples are mostly from the medical and bio-medical area as a consequence
- We are exploring potential of change in the SSH in



## → Much more than catalogues and discovery systems: Visualise and Explore Cultural Context

- The graph of writers and thinkers and how they are connected:

Version 1 (is this one really useful?)

Version 2

A more dynamic rendering: Relfinder

- Or a non-linear reading version of this presentation
- Or again a Finnish example (Kulttuurisampo):

<http://www.kulttuurisampo.fi/kulsa/historiallisetKartat.shtml>

- Or finally Obama vs. Palin:

[http://truthy.indiana.edu/memedetail?id=324&resmin=45&theme\\_id=4](http://truthy.indiana.edu/memedetail?id=324&resmin=45&theme_id=4) vs.

[http://truthy.indiana.edu/memedetail?id=783&resmin=45&theme\\_id=4](http://truthy.indiana.edu/memedetail?id=783&resmin=45&theme_id=4)



# An Opportunity for (Research) Libraries ... ... and what it needs to do to be up to it



# “What do you do with a million books?”

## (Greg Crane)

- Digitisation and semantic publishing result in
  - growing quantity
  - increased complexity
- Well beyond scholarly processing capacity (=reading faculty)
- Scientists and Scholars will badly need help in three areas:
  - **Semantic abstracting, named entity recognition** for “strategic reading” (Renear)
  - **Contextualisation** of information objects
  - Robust **reasoning and inferencing** yielding digital heuristics



⇒ **Opportunities for Research Libraries!**

Librarian Agendas at the End of the Gutenberg Paranthesis

Stefan Gradmann, ExLibris National Library Meeting, Singapore, 21/08/2013



# Ceci n'est pas une bibliothèque

KU LEUVEN





# Ceci n'est pas une bibliothèque





# Catalogue



The card catalog in the nave of Sterling Memorial Library at Yale University. Picture by Henry Trotter, 2005.



# Catalogue Entry: MARC Record

**LEADER:** 00495cam 2200169 a 4500  
001 94205672  
003 DLC  
005 2004010812243.0  
008 940907s1926 gw 000 1 ger  
010 \$a 94205672  
040 \$aDLC\$cDLC\$dDLC  
050 00 \$aPT2621.A26\$bS36 1926  
100 1 \$aKafka, Franz,\$d1883-1924  
245 14 \$aDas Schloss :\$bRoman /\$cFranz Kafka.  
260 \$aMünchen :\$bKurt wolff verlag,\$c1926.  
300 \$a505 p. ;\$c20 cm.  
561 \$aSource: Bequest of Aaron Copland, Fall, 1989.\$5DLC

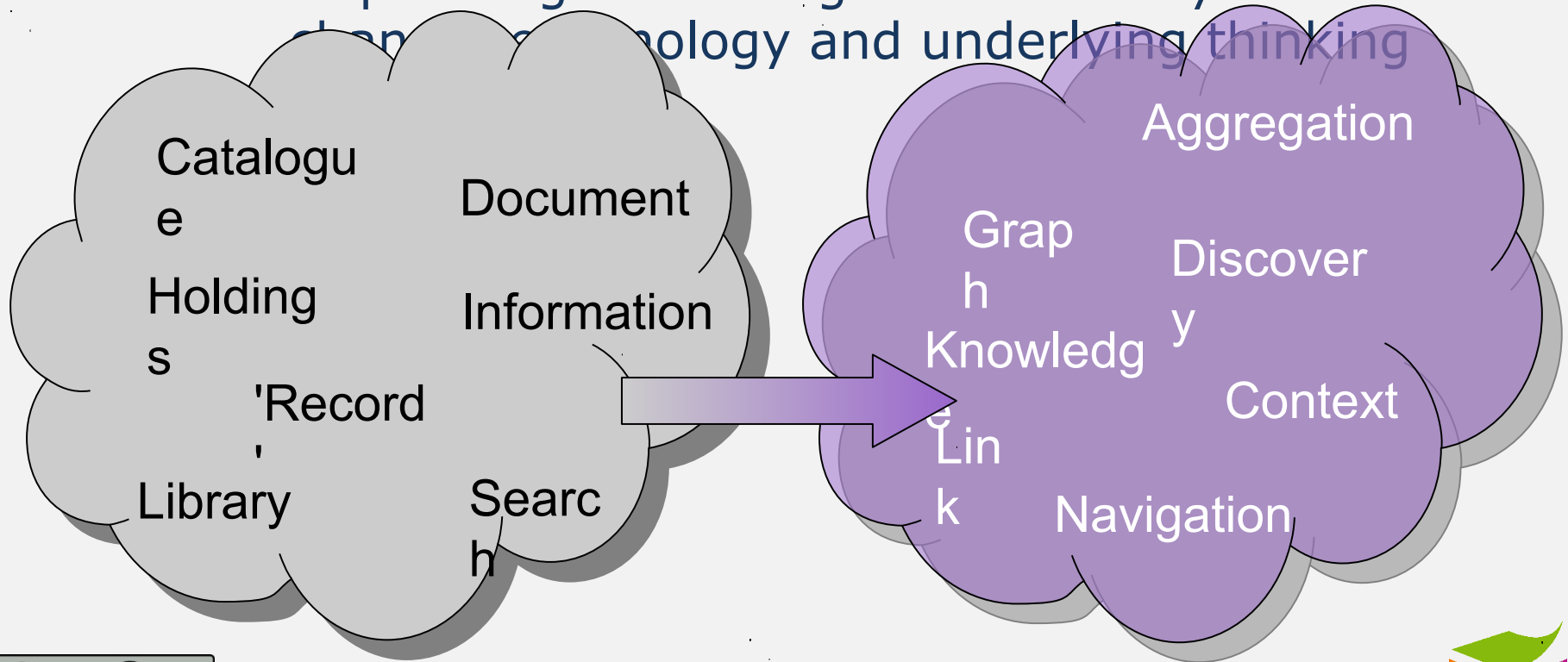


# 'Library Collections'



# Change Thinking, Change Terminology!

- Libraries will serve research as part of the Linked Open Data web – or else risk becoming insignificant.
- For operating this change we definitely need to change terminology and underlying thinking



# Sticking to empty metaphors ...

"What's in a name? That which we call a rose  
By any other name would smell as sweet."

(Shakespeare, Romeo and Juliet (II, ii, 1-2))

Why then do we stick to emptied metaphors?

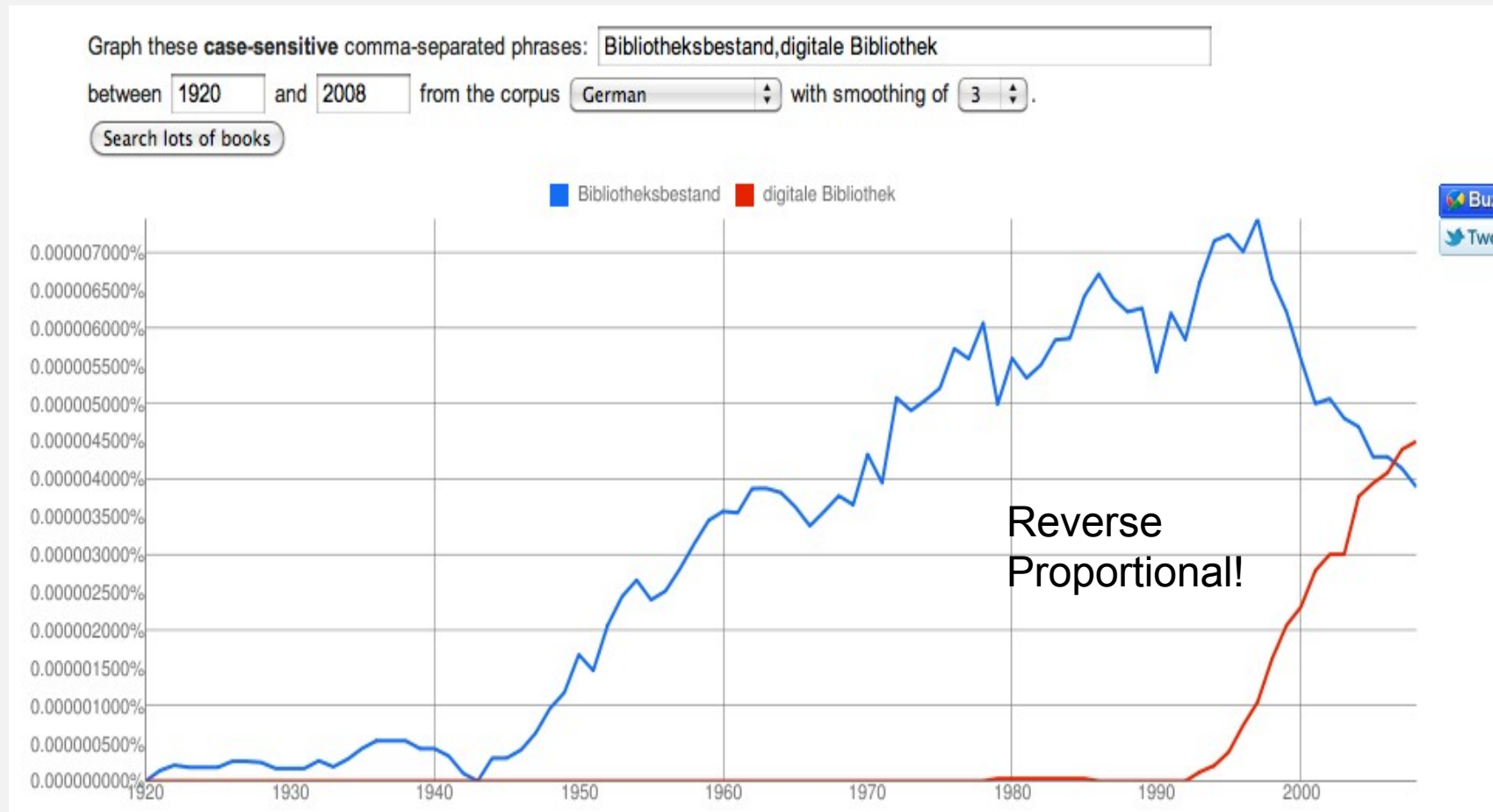
- ... because they **constitute identity** (a very bad reason!)
- ... because they guarantee **institutional persistency** (a fallacy!)
- ... because we are afraid of substantial changes and believe in **things changing only once we use new terms** (dangerously childish!)
- ... or simply because **we do not have new terms** yet?

**Let us then start looking for them!**



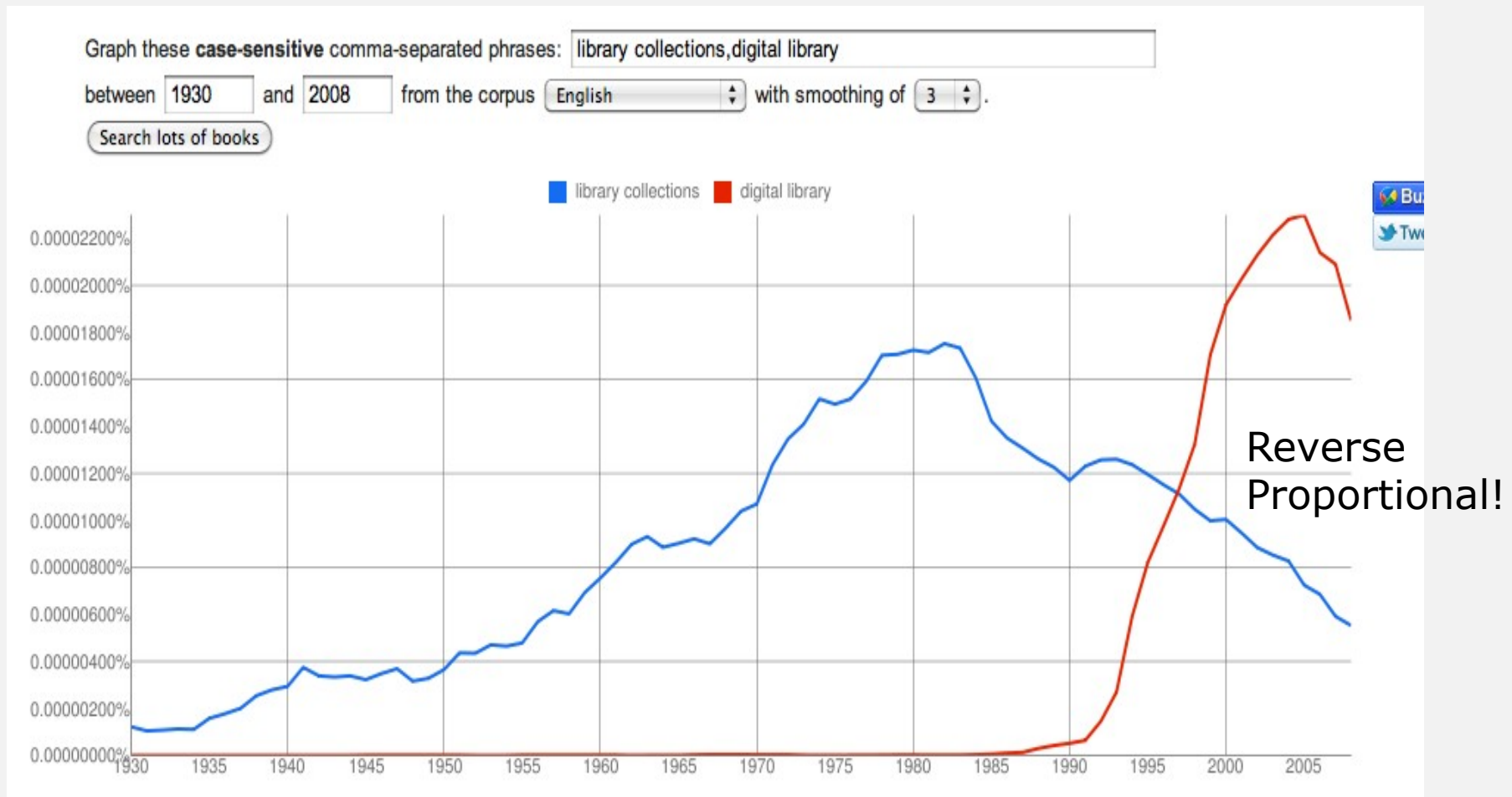


# From 'Catalogues' to 'Graphs': old terms – new terms (1)

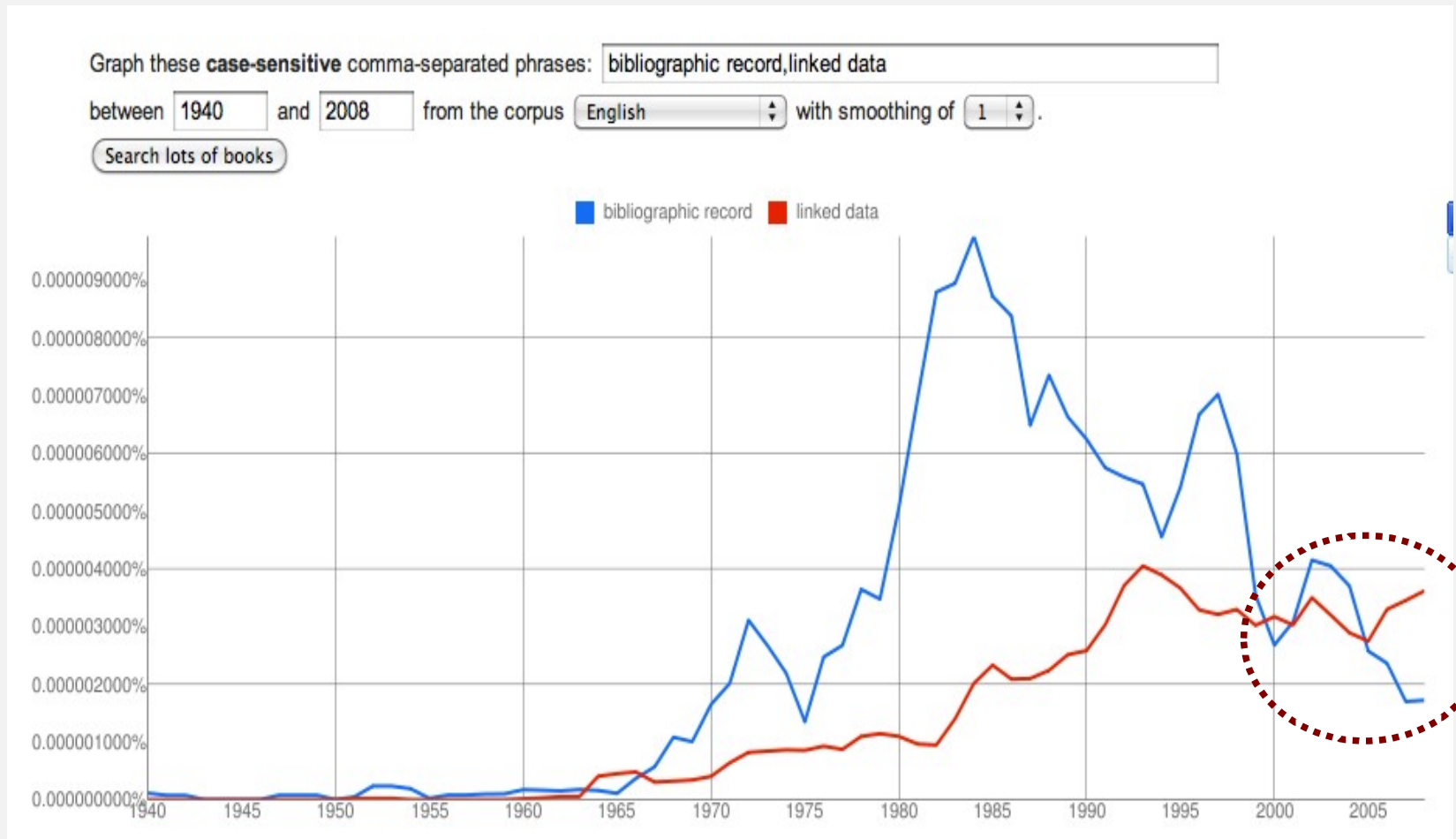




# From 'Catalogues' to 'Graphs': old terms – new terms (2)



# From 'Catalogues' to 'Graphs': old terms – new terms (3)



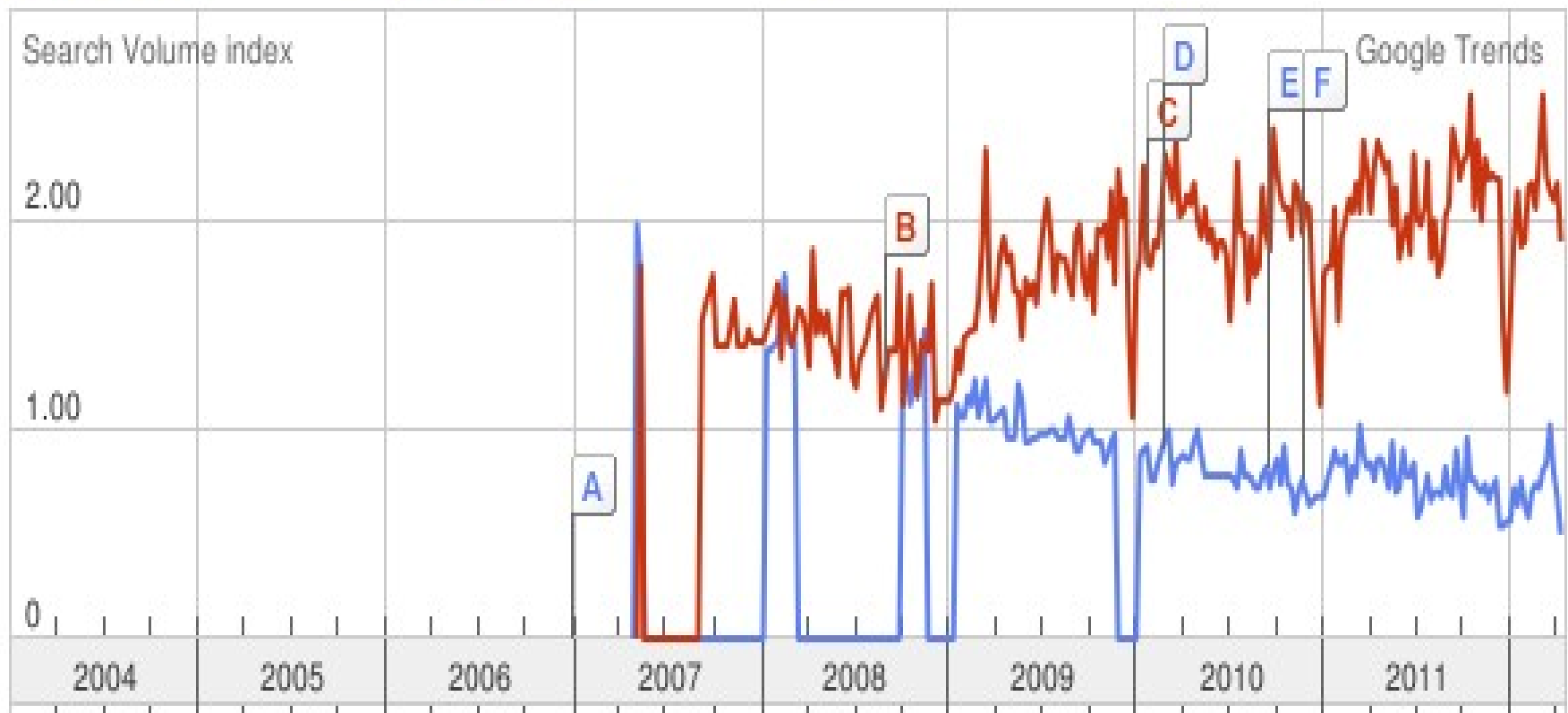
# From 'Catalogues' to 'Graphs': old terms – new terms (4)

cataloguing

0.38

linked data

1.00



# A Commonwealth of Data and Publications ...

... do we want it?

... do we want to be part of it??



# Strategy Issues: two questions and two recommendations

- Opportunities for Research Libraries ...
  - ... certainly – but for how many of them?
  - ... and what about the ones left behind?
- And what about the National Libraries?
  - Are you not the least concerned?
  - Hmm ... think about digital legal deposit and remember slide 26
- Make sure your staff knows what RDF is and what can be built on it
  - And use the introduction of RDA as a stepping stone!
- Proactively reposition your institutions ...
  - ... as close to scholarly research as possible
  - ... before they are repositioned by others



# Suggested Reading

Gregory Crane (2006): What Do you Do with a Million Books? In: Dlib Magazine, Vol. 12, March. (<http://bit.ly/JhzF90>)

Gutenberg Paranthesis Research Group / University of Southern Denmark: Position Paper (<http://bit.ly/JjGKb6>)

Barend Mons, Jan Velterop: Nano-Publication in the e-science era (<http://bit.ly/IISMGt>)

David Parry: Burn the Boats/Books. Presentation to Digital Writing and Research Lab, Austin. (<http://bit.ly/JYLIJv>)

Alan Renear, Carol Palmer (2009): Strategic Reading, Ontologies and the Future of scientific Publishing. In: Science, August 2009, p. 828 – 832.

David Shotton (2009a): Semantic Publishing. The coming revolution in scientific journal publishing. Learned Publishing Volume 22, No 2, 85–94, April 2009; doi:10.1087/2009202

David Shotton et al. (2009b): Adventures in Semantic Publishing: Exemplar Semantic Enhancements of a Research Article (<http://bit.ly/IgT5Km>)

Stiegler, B. (2012). Die Aufklärung in the Age of Philosophical Engineering. Computational Culture. (<http://bit.ly/14VR5ph>)





**Thank you for your patience and attention**

**Questions (en)?**

**Vragen (nl)?**

**Fragen (de)?**

**Questions (fr)?**

